Introduction:
I currently have a friend in Toronto who lives in the Rosedale neighborhood. They have opened a successful business (yoga studio) in this area. They are currently analyzing expanding the business and opening several new locations. They feel that the biggest advantage that they have had in successfully running their business has been the lack of competition around them. Given that Toronto is a very large city, it would impossible to know every single neighborhood, let alone the potential spaces that a business could be opened in. In order to do this my friend has approached me to narrow down the search area and perhaps even provide a best recommendation for other neighborhoods to possibly open another location in. The end deliverable was designed to be a list of neighborhoods similar to Rosedale.

Data:
In order to solve this problem we will first need to figure out what neighborhoods are similar to Rosedale since the business owner feels this is the largest contributor to their success. In order to do this we will use Foursquare mapping data in order to pull venues from the neighborhoods. There are many features that you can extract from this foursquare data. These include things like venues, comments, photos, etc. In order to accomplish this task we will simply be using the venue and the categories of the venues. We will also be using data from Wikipedia in order to visualize the data.
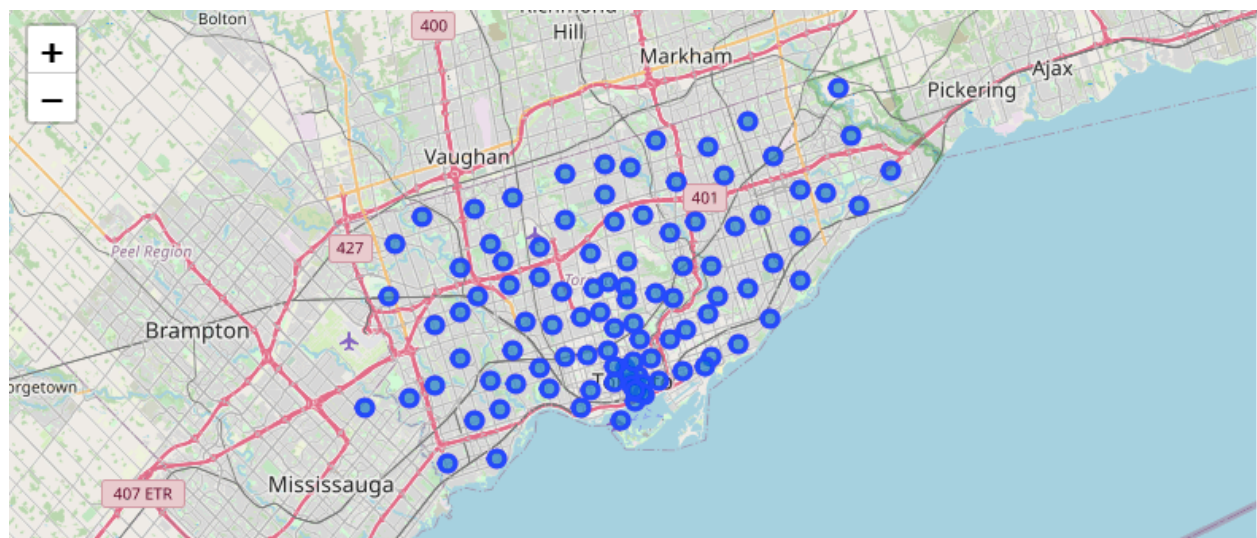

Methodology:
The exploratory data analysis entailed scraping the list of neighborhoods from the Toronto Wikipedia neighborhood list. In order to ensure that this was done correctly we pulled the latitude and longitudes as well in order to map the neighborhoods. We also made sure there was no NaN in the data, ensuring it was clean. After this was done we pulled the mapping data from Foursquare. This was done for all the neighborhoods in Toronto. In order to determine which neighborhood was the most similar to Rosedale we used k-means clustering. To do this we pulled the venues from the neighborhood and then used frequency distributions to

determine the characteristics of every neighborhood. Once this was done the k-means clustering was applied. At this point we were able to pull a list of similar neighborhoods to Rosedale. This would probably be what the client would receive but we then did another iteration of k-means clustering in order to identify the most similar neighborhood to Rosedale in case the client only wanted one neighborhood to open a second business.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	
0	Agincourt	Latin American Restaurant	Lounge	Breakfast Spot	Chinese Restaurant	
1	Alderwood / Long Branch	Pizza Place	Sandwich Place	Skating Rink	Gym	
2	Bathurst Manor / Wilson Heights / Downsview North	Coffee Shop	Bank	Grocery Store	Fried Chicken Joint	
3	Bayview Village	Café	Bank	Chinese Restaurant	Japanese Restaurant	
4	Bedford Park / Lawrence Manor East	Coffee Shop	Italian Restaurant	Restaurant	Sandwich Place	
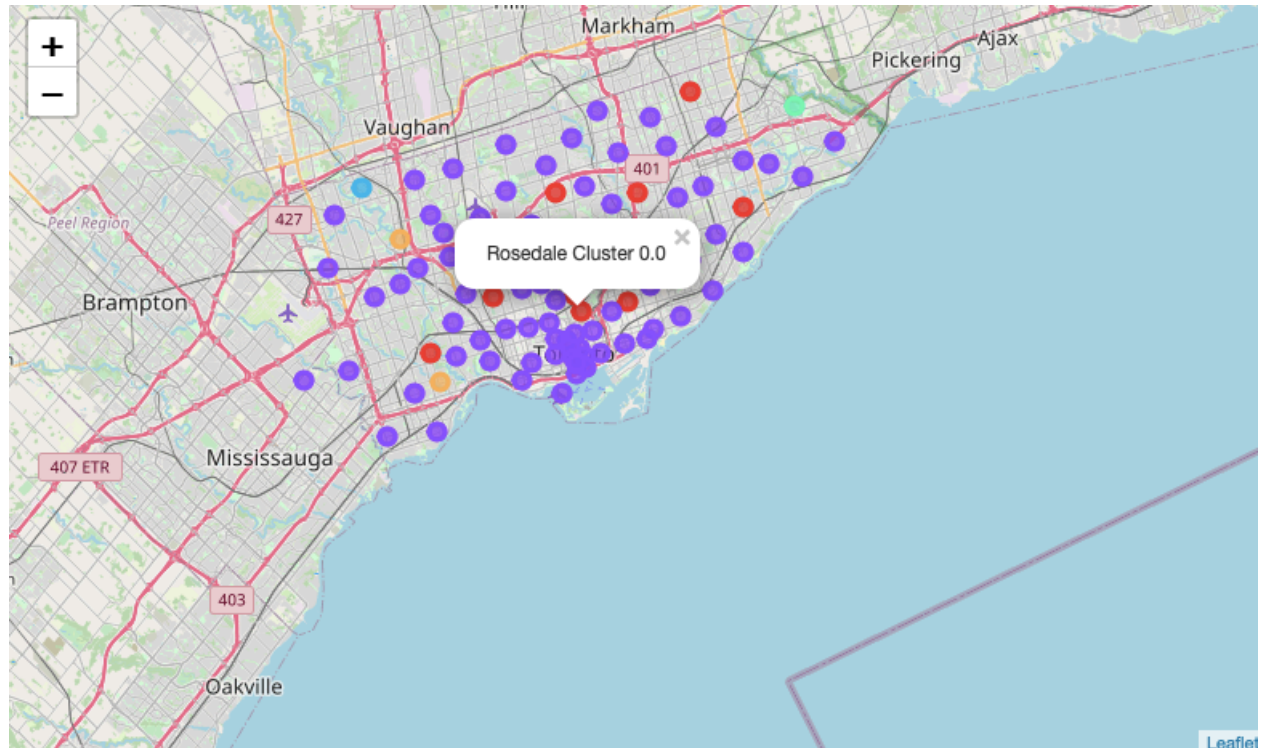
Results:

The most similar neighborhood to Rosedale is Milliken / Agincourt North / Steeles East / L'Amoreaux East. That said the more appropriate data would probably the 9 similar neighborhoods to Rosedale based on the clustering results.

	Neighborhood
0	Parkwoods
1	Caledonia-Fairbanks
2	Scarborough Village
3	East Toronto
4	Lawrence Park
5	York Mills West
6	Moore Park / Summerhill East
7	Milliken / Agincourt North / Steeles East / L'...
8	Rosedale
9	The Kingsway / Montgomery Road / Old Mill North

```
The best neighborhood to open a second a location is Milliken / Agincourt North / Steeles Eas
t / L'Amoreaux East
```

Discussion:
At this point the client would probably want to start looking at possible real estate locations for their new business in these areas. We have successfully taken a list of 93 neighborhoods and narrowed the search area to 9. That said, if the client wishes to focus on one area Milliken would be the best. Given the smaller sample size used in the second clustering I would advise the client to be cautious of only taking the one neighborhood. If they decided that they did want us to find the best neighborhood and not just a list, I would then offer to go back and calculate other distances or similarity scores which may supplement the second round of k-means clustering that we did.

Conclusion:
The client's stated goal of finding a list of neighborhoods that are similar to Rosedale was accomplished using K-means clustering. This was done using Foursquare data that allowed us to characterize the neighborhoods based on the type and numbers of venues that are in each neighborhood. This resulted in a list of neighborhoods where the client can focus on in order to open a second location of their business. We have successfully narrowed down the search for a second possible location from over 90 neighborhoods to 9.